IN THE CLAIMS:

All pending claims and their present status are produced below.

1.   (Previously Presented) An apparatus for direct annotation of objects, the apparatus comprising:

a display device for displaying one or more images;

an audio input device for receiving an audio signal;

a storage device for storing a plurality of different visual notations each comprising a text or a graphic image and for storing a plurality of corresponding audio signals;

a direct annotation creation module coupled to receive the audio signal from the audio input device and to receive a reference to a location within an image on the display device, the direct annotation creation module, in response to receiving the audio signal and the reference to the location within the image, automatically creating an annotation object, independent from the image, that associates the input audio signal, the location and one of the plurality of different visual notations; and

an audio vocabulary comparison module coupled to the audio input device, the storage device and the direct annotation creation module, the audio vocabulary comparison module receiving audio input and finding a corresponding one of the plurality of different visual notations that matches the audio input.

2. (Original) The apparatus of claim 1 further comprising an annotation display module coupled to the direct annotation creation module, the annotation display module generating symbols or text representing the annotation objects.

3. (Original) The apparatus of claim 1 further comprising an annotation audio output module coupled to the direct annotation creation module, the annotation audio output module generating audio output in response to user selection of an annotation symbol representing an annotation object.

4. (Canceled).

5. (Previously Presented) The apparatus of claim 1 further comprising:
an audio vocabulary storage for storing a plurality of audio signals and corresponding text strings;
a dynamic vocabulary updating module coupled to the audio vocabulary storage and the audio input device, the dynamic vocabulary updating module for displaying an interface to create a new entry in the audio vocabulary storage, the dynamic vocabulary updating module receiving an audio input and a text string and creating the new entry in the audio vocabulary storage that includes a new visual annotation.

6. (Original) The apparatus of claim 1 further comprising a media object cache for storing media and annotation objects.

7.   (Canceled).

8.   (Canceled).

9.   (Previously Presented) An apparatus for direct annotation of objects, the apparatus comprising:

a media object storage for storing media, annotation objects, a plurality of different visual notations each comprising a text or a graphic image and a plurality of corresponding audio signals;

a direct annotation creation module coupled to receive an audio signal, a selected visual notation from the plurality of different visual notations and a reference to a location within an image, the direct annotation creation module, in response to receiving the audio signal or the reference to the location within the image, automatically creating an annotation object, independent of the image, that associates the audio signal, the selected visual notation and the location, and the direct annotation creation module storing the audio annotation in the media object storage;

an audio vocabulary comparison module coupled to the media object storage and the direct annotation creation module, the audio vocabulary comparison module receiving audio input and finding a corresponding one of the plurality of different visual notations that matches the audio signal; and

an annotation output module coupled to the direct annotation creation module, the annotation output module generating audio or visual output in response to user selection of an annotation symbol representing the annotation object.

10. (Canceled)

11. (Previously Presented) The method of claim 26, wherein the step of displaying is performed before or simultaneously with the step of receiving.

12. (Previously Presented) The method of claim 26, wherein the step of receiving is performed before or simultaneously with the step of displaying.

13. (Canceled)

14. (Previously Presented) The method of claim 26, further comprising the step of displaying the one of the plurality of different visual notations to indicate that the image has an annotation.

15. (Canceled)

16. (Previously Presented) The method of claim 26, wherein the step of creating an annotation object includes storing the annotation object in an object storage.

17. (Canceled).

18. (Canceled).

19. (Canceled).

20.   (Canceled).

21.   (Canceled).

22.   (Canceled).

23.   (Canceled).

24.   (Canceled).

25.   (Canceled).

26. (Previously Presented) A computer implemented method for direct annotation of
objects, the method comprising the steps of:

displaying an image;

receiving audio input;

detecting selection of a location within the image;

comparing the audio input to a vocabulary to produce text or a graphic image; and

creating an annotation object, independent of the selected image, that provides an

association between the image, the audio input, the selected location, one of a

plurality of different visual notations comprising text or a graphic image, the

annotation object including at least a text annotation field, an image reference

field, and an annotation location field, the creating step occurring

automatically in response to the receiving or detecting.

27. (Original) The method of claim 26, further comprising the step of recording the audio input received.

28. (Previously Presented) The method of claim 27, wherein the step of creating the annotation object includes creating an annotation object including a reference to the selected location, the recorded audio input and one of the plurality of different visual annotations, and storing the annotation object in an object storage.

29. (Previously Presented) The method of claim 26, wherein the step of creating an annotation object includes storing the text as part of the annotation object.

30. (Original) The method of claim 26, further comprising the steps of:

determining if the audio input has a matching entry in the vocabulary; and

storing the entry as part of the annotation object if the audio input has a matching

      entry in the vocabulary.

31. (Previously Presented) The method of claim 30, further comprising the steps of:

determining if the audio input has a close match in the vocabulary;

displaying the close matches;

receiving input selecting a close match; and

storing the selected close match as part of the annotation object if the audio input has

      a close match in the vocabulary.

32. (Previously Presented) The method of claim 31, further comprising the step of displaying a message that the image has not been annotated if there is neither a matching entry in the vocabulary nor a close match in the vocabulary.

33. (Previously Presented) The method of claim 31, further comprising the following steps if there is neither a matching entry in the vocabulary nor a close match in the vocabulary:

receiving text input corresponding to the audio input;

updating the vocabulary with a new entry including the audio input and the text input;

and

wherein the received text is stored as part of the annotation object.

34. (Previously Presented) The method of claim 26, further comprising the steps of:

receiving text input corresponding to the audio input;

updating the vocabulary with a new entry including the audio input and the text input.

35. (Previously Presented) A computer implemented method for displaying objects with annotations, the method comprising the steps of:

retrieving an image;

displaying the image with one of a plurality of different visual notations to indicate

that an annotation exists;

receiving user selection of the one visual notation;

generating the annotation automatically, in response to user input of a location within

the image and an audio input;

outputting the annotation associated with the selected visual notation;

determining whether the annotation includes text;

retrieving a text annotation for the selected visual notation; and

displaying the retrieved text with the image.

36.   (Previously Presented) The method of claim 35, wherein the annotation is text and the step of outputting is displaying the text proximate the image that it annotates.

37.   (Previously Presented) The method of claim 35, wherein the annotation is an audio signal and the step of outputting is playing the audio signal.

38.   (Canceled)

39.   (Previously Presented) The method of claim 35, further comprising the steps of:

determining whether the annotation includes an audio signal;

retrieving an audio signal for the selected visual annotation; and

wherein the step of outputting is playing the audio signal.

40.   (Previously Presented) A computer implemented method for retrieving images, the method comprising the steps of:

receiving audio input;

determining annotation objects that reference a close match to the audio input, each

annotation object generated automatically in response to user input of a

location within an image and an audio signal, where a recording of the audio

signal is terminated automatically based on a predetermined audio level;

retrieving the images that are referenced by the determined annotation objects; and

displaying the retrieved images, one of a plurality of different visual notations for the

annotation object and wherein the annotation object includes at least an audio

input field, an image reference field, and an annotation location field.

41.   (Previously Presented) The method of claim 40, wherein the step of

determining annotation objects further comprises the steps of:

comparing the audio input to an audio signal reference of the annotation object; and

determining a close match between the audio input and the audio signal reference of

the annotation object if a probability metric is greater than a threshold of 80%.

42.   (Previously Presented) The method of claim 40, wherein the step of

determining annotation objects further comprises the steps of:

determining the annotation objects for a plurality of images;

for each annotation object, comparing the audio input to an audio signal reference of

the annotation object; and

determining a close match between the audio input and the audio signal reference of

the annotation object if a probability metric is greater than an a threshold of

80%.